What Makes Expectations Great?
MDPs with Bonuses in the Bargain

J. Michael Steele [1]

Wharton School
University of Pennsylvania

April 6, 2017

[1] including joint work with A. Arlotto (Duke) and N. Gans (Wharton)

1. Some Worrisome Questions with Useful Answers
2. Three motivating examples
3. Rich Class of MDPs where Means Bound Variances
4. Beyond Means-Bound-Variances: Sequential Knapsack Problem
5. Markov Additive Functionals of Non-Homogenous MCs — with a Twist
6. Motivating examples: finite-horizon Markov decision problems
7. Dobrushin's CLT and failure of state-space enlargement
8. MDP Focused Central Limit Theorem
9. Conclusions

What About Expectations and MDPs Were You Afraid to Ask?

- Sequential decision problems: The tool of choice is almost always dynamic programming and the objective is almost always maximization of an expected reward (or minimization of an expected cost).

- In practice, we know dynamic programming often performs admirably. Still, sooner or later, the little voice in our head cannot help but ask, "What about the Saint Petersburg Paradox?"

- Nub of the Problem: The realized reward of the decision maker is a random variable, and, for all we know *a priori*, our myopic focus on means might be disastrous.

- Is there an analytical basis for the sensible performance of mean-focused dynamic programming?

- Our first goal is to isolate a rich class of dynamic programs were the mean-focused optimality also guarantees low variability. This is the bonus to our bargain.

- Beyond variance bounds, there is the richer goal of limit theorems for the realized total reward. In MDPs there are many sources of dependence and their participation is typically non-linear. This creates a challenging long term agenda. Some concrete progress has been made.

1. Some Worrisome Questions with Useful Answers
2. Three motivating examples
3. Rich Class of MDPs where Means Bound Variances
4. Beyond Means-Bound-Variances: Sequential Knapsack Problem
5. Markov Additive Functionals of Non-Homogenous MCs — with a Twist
6. Motivating examples: finite-horizon Markov decision problems
7. Dobrushin's CLT and failure of state-space enlargement
8. MDP Focused Central Limit Theorem
9. Conclusions

Three motivating examples: one common question. . .

Example one: sequential knapsack problem (Coffman et al., 1987)

- Knapsack capacity $c \in (0, \infty)$
- Item sizes: $Y_1, Y_2, \ldots, Y_n$ independent, continuous distribution F
- Decision: Viewing Y_t, $1 \leq t \leq n$, sequentially; decide *include/exclude*
- Knapsack policy π_n: the number of items included is

$$R_n(\pi_n) = \max \left\{ k : \sum_{i=1}^{k} Y_{\tau_i} \leq c \right\},$$

τ_i, index of the *i*th item included

- Objective: $\sup_{\pi_n} \mathbb{E}\left[R_n(\pi_n)\right]$
- π_n^*: optimal Markov deterministic policy
- Basic Question: What can you say about

$$\mathrm{Var}\left[R_n(\pi_n^*)\right] \ \textbf{?}$$

Three motivating examples: one common question...

Example two: quantity-based revenue management (Talluri and van Ryzin, 2004)

- Total capacity $c \in \mathbb{N}$
- Prices: $Y_1, Y_2, \ldots, Y_n$ independent, $Y_t \in \{y_0, \ldots, y_\eta\} \overset{\text{dist}}{\sim} \{p_0(t), \ldots, p_\eta(t)\}$
- Decision: *sell/do not sell* one unit of capacity at price Y_t
- Selling policy π_n: total revenues are

$$R_n(\pi_n) = \max\left\{\sum_{i=1}^{k} Y_{\tau_i} \, : \; k \le c\right\},$$

  τ_i, index of the ith unit of capacity sold
- Objective: $\sup_{\pi_n} \mathbb{E}\left[R_n(\pi_n)\right]$
- π_n^*: optimal Markov deterministic policy
- Basic Question: What can you say about

$$\mathrm{Var}\left[R_n(\pi_n^*)\right] \;\; \textbf{?}$$

Three motivating examples: one common question. . .

Example three: sequential investment problem (Samuelson, 1969; Derman et al., 1975; Prastacos, 1983)

- Capital available $c \in (0, \infty)$
- Investment opportunities: $Y_1, \ldots, Y_n$, independent, known distribution F
- Decision: how much to invest in each opportunity
- Investment policy π_n: total return is

$$R_n(\pi_n) = \sum_{t=1}^{n} r(Y_t, A_t),$$

  where $r(y, a)$ is the return of investing a units of capital when $Y_t = y$

- Objective: $\sup_{\pi_n} \mathbb{E}\left[R_n(\pi_n)\right]$
- π_n^*: optimal Markov deterministic policy
- Basic Question: What can you say about

$$\mathrm{Var}\left[R_n(\pi_n^*)\right] \ \textbf{?}$$

What Do We Gain From Understanding $\mathrm{Var}\,[R_n(\pi_n^*)]$?

- Without some understanding of variability, the decision maker is simply set adrift. Consider the story of the “reasonable” house seller.
- In practical contexts, one can fall back on simulation studies, but inevitably one is left with uncertainties of several flavors. For example, what model tweaks suffice to give well behaved variance bounds?
- More ambitiously, we would hope to have some precise understanding of the distribution of the realized reward.
- For example, it’s often natural to expect the realize reward to be asymptotically normally distributed. Can one have any hope of such a refined understanding without first understanding the behavior of the variance?

Good Behavior of the Example MDPs

Theorem (Arlotto, Gans and S., OR 2014)

The optimal total reward $R_n(\pi_n^*)$ of the knapsack, revenue management and investment problems satisfies

$$\mathrm{Var}\,[R_n(\pi_n^*)] \le K\,\mathbb{E}\,[R_n(\pi_n^*)] \quad \textit{for each } 1 \le n < \infty,$$

where

- $K = 1$ *in the sequential knapsack problem*
- $K = \max\{y_0, \dots, y_\eta\}$, *the largest price in the revenue management problem*
- $K = \sup\{r(y,a)\}$ *in the investment problem*

HEADS-UP. The *means-bound-variance relations* are of particular interest in those problems where the expectation grows sub-linearly. For example, in the knapsack problem for random variables with bounded support, the mean is asymptotic to $c\sqrt{n}$.

Variance Bounds: Simplest Implications

Coverage Intervals for Realized Rewards

For $\alpha > 1$,

$$-\alpha\sqrt{K\,\mathbb{E}\,[R_n(\pi_n^*)]} + \mathbb{E}\,[R_n(\pi_n^*)] \leq R_n(\pi_n^*) \leq \mathbb{E}\,[R_n(\pi_n^*)] + \alpha\sqrt{K\,\mathbb{E}\,[R_n(\pi_n^*)]}$$

with probability at least $1 - 1/\alpha^2$

Small Coefficient of Variation

The optimal total reward has relatively small coefficient of variation:

$$\text{CoeffVar}[R_n(\pi_n^*)] = \frac{\sqrt{\text{Var}\,[R_n(\pi_n^*)]}}{\mathbb{E}[R_n(\pi_n^*)]} \leq \sqrt{\frac{K}{\mathbb{E}[R_n(\pi_n^*)]}}$$

Weak Law of Large Numbers for the Optimal Total Reward

If $\mathbb{E}\,[R_n(\pi_n^*)] \to \infty$ as $n \to \infty$ then

$$\frac{R_n(\pi_n^*)}{\mathbb{E}[R_n(\pi_n^*)]} \xrightarrow{p} 1$$

1. Some Worrisome Questions with Useful Answers
2. Three motivating examples
3. Rich Class of MDPs where Means Bound Variances
4. Beyond Means-Bound-Variances: Sequential Knapsack Problem
5. Markov Additive Functionals of Non-Homogenous MCs — with a Twist
6. Motivating examples: finite-horizon Markov decision problems
7. Dobrushin's CLT and failure of state-space enlargement
8. MDP Focused Central Limit Theorem
9. Conclusions

General MDP Framework

$$\left(\mathcal{X}, \mathcal{Y}, \mathcal{A}, f, r, n \right)$$

- $\mathcal{X}$ is the state space; at each t the DM knows the state of the system $x \in \mathcal{X}$
  - Knapsack example: x is the remaining capacity
- The independent sequence $Y_1, Y_2, \ldots Y_n$ takes value in $\mathcal{Y}$
  - Knapsack example: $y \in \mathcal{Y}$ is the size of the item that is presented
- Action space: $\mathcal{A}(t, x, y) \subseteq \mathcal{A}$ is the set of admissible actions for (x, y) at t
  - Knapsack example: “select”; “do not select”
- State transition function: $f(t, x, y, a)$ state that one reaches for $a \in \mathcal{A}(t, x, y)$
  - Knapsack example: $f(t, x, y, \mathrm{select}) = \mathrm{x} - \mathrm{y}$; $f(t, x, y, \mathrm{do\,not\,select}) = \mathrm{x}$
- Reward function: $r(t, x, y, a)$ reward for taking action a at time t when at (x, y)
  - Knapsack example: $r(t, x, y, \mathrm{select}) = 1$; $r(t, x, y, \mathrm{do\,not\,select}) = 0$
- Time horizon: $n < \infty$

MDPs where Means Bound Variances

- $\Pi(n)$ set of all feasible Markov deterministic policies for the n-period problem

- Reward of policy π up to time k

$$R_k(\pi) = \sum_{t=1}^{k} r(t, X_t, Y_t, A_t), \qquad X_1 = \bar{x}, \quad 1 \leq k \leq n$$

- Expected total reward criterion, i.e. we are looking for $\pi_n^* \in \Pi(n)$ such that

$$\mathbb{E}[R_n(\pi_n^*)] = \sup_{\pi \in \Pi(n)} \mathbb{E}[R_n(\pi)].$$

- Bellman equation: for each $1 \leq t \leq n$ and for $x \in \mathcal{X}$,

$$v_t(x) = \mathbb{E}\left[\sup_{a \in \mathcal{A}(t,x,Y_t)} \{ r(t, x, Y_t, a) + v_{t+1}\left(f(t, x, Y_t, a)\right)\} \right],$$

  - $v_{n+1}(x) = 0$ for all $x \in \mathcal{X}$, and
  - $v_1(\bar{x}) = \mathbb{E}[R_n(\pi_n^*)]$

Three Basic Properties

Property (Non-negative and Bounded Rewards)

There is a constant $K < \infty$ *such that* $0 \leq r(t, x, y, a) \leq K$ *for all triples* (x, y, a) *and all times* $1 \leq t \leq n$.

Property (Existence of a Do-nothing Action)

For each time $1 \leq t \leq n$ *and pair* (x, y)*, the set of actions* $\mathcal{A}(t, x, y)$ *includes a do-nothing action* a^0 *such that*

$$f(t, x, y, a^0) = x$$

Property (Optimal Action Monotonicity)

For each time $1 \leq t \leq n$ *and state* $x \in \mathcal{X}$ *one has the inequality*

$$v_{t+1}(x^*) \leq v_{t+1}(x)$$

for all $x^* = f(t, x, y, a^*)$ *for some* $y \in \mathcal{Y}$ *and any optimal action* $a^* \in \mathcal{A}(t, x, y)$.

MDPs where Means Bound Variances

Theorem (Arlotto, Gans, S. *OR* 2014)

Suppose that the Markov decision problem $(\mathcal{X}, \mathcal{Y}, \mathcal{A}, f, r, n)$ *satisfies reward non-negativity and boundedness, existence of a do-nothing action and optimal action monotonicity. If* $\pi_n^* \in \Pi(n)$ *is any Markov deterministic policy such that*

$$\mathbb{E}[R_n(\pi_n^*)] = \sup_{\pi \in \Pi(n)} \mathbb{E}[R_n(\pi)],$$

then

$$\mathrm{Var}[R_n(\pi_n^*)] \leq K\, \mathbb{E}[R_n(\pi_n^*)],$$

where K *is the uniform bound on the one-period reward function.*

Some Implications:

- Well-defined coverage intervals
- Small Coefficient of Variation
- Weak law of large numbers for the optimal total reward
- Ameliorated Saint Petersburg Paradox Anxiety (but ...)

Understanding the Three Crucial Properties

In summary, the variance of the optimal total reward is "small" provided that we have the three conditions: . . .

Property 1: Non-negative and bounded rewards

Property 2: Existence of a do-nothing action

Property 3: Optimal action monotonicity, $v_{t+1}(x^*) \leq v_{t+1}(x)$

Are these easy to check? Fortunately, the answer is "Yes."

Optimal action monotonicity: sufficient conditions

Sufficient Conditions

A Markov decision problem $(\mathcal{X}, \mathcal{Y}, \mathcal{A}, f, r, n)$ *satisfies optimal action monotonicity if:*

(i) *the state space* $\mathcal{X}$ *is a subset of a finite-dimensional Euclidean space equipped with a partial order* $\preceq$*;*

(ii) *for each* $y \in \mathcal{Y}$, $1 \leq t \leq n$ *and each optimal action* $a^* \in \mathcal{A}(t, x, y)$ *one has* $f(t, x, y, a^*) \equiv x^* \preceq x$

(iii) *for each* $1 \leq t \leq n$*, the map* $x \mapsto v_t(x)$ *is non-decreasing: i.e.* $x \preceq x'$ *implies* $v_t(x) \leq v_t(x')$*;*

Remark

Analogously, one can require that

(ii) $x \preceq x^* \equiv f(t, x, y, a^*)$*;*

(iii) *the map* $x \mapsto v_t(x)$ *is non-increasing: i.e.* $x \preceq x'$ *implies* $v_t(x') \leq v_t(x)$*.*

Further Examples

Naturally there are MDPs that fail to have one or more of the basic properties used here, but, there is a robust supply of of MDPs where we do have (1) reward non-negativity and boundedness, (2) existence of a do-nothing action, and (3) optimal action monotonicity. For example we have:

- General dynamic and stochastic knapsack formulations (Papastavrou, Rajagopalan, and Kleywegt, 1996)
- Network capacity control problems in revenue management (Talluri and van Ryzin, 2004)
- Combinatorial optimization: sequential selection of monotone, unimodal and d-modal subsequences (Arlotto and S., 2011)
- **Your favorite MDP!** Please add to the list.

Two Variations on the MDP Framework

There are two basic variations on the basic MDP Framework, where one can again obtain the means-bound-variances inequality. In many contexts one has discounting or additional post-action randomness, and these can be accommodated without much difficulty. Specifically, one can deal with

- Finite horizon discounted MDPs
- MDPs with within-period uncertainty that realizes after the decision maker chooses the optimal action:
  - Such uncertainty might affect both the optimal one-period reward and the optimal state-transition
  - An important example that belongs to this class are stochastic depletion problems (Chan and Farias, 2009)

1. Some Worrisome Questions with Useful Answers
2. Three motivating examples
3. Rich Class of MDPs where Means Bound Variances
4. Beyond Means-Bound-Variances: Sequential Knapsack Problem
5. Markov Additive Functionals of Non-Homogenous MCs — with a Twist
6. Motivating examples: finite-horizon Markov decision problems
7. Dobrushin's CLT and failure of state-space enlargement
8. MDP Focused Central Limit Theorem
9. Conclusions

Sharper Focus: The Simplest Sequential Knapsack Problem

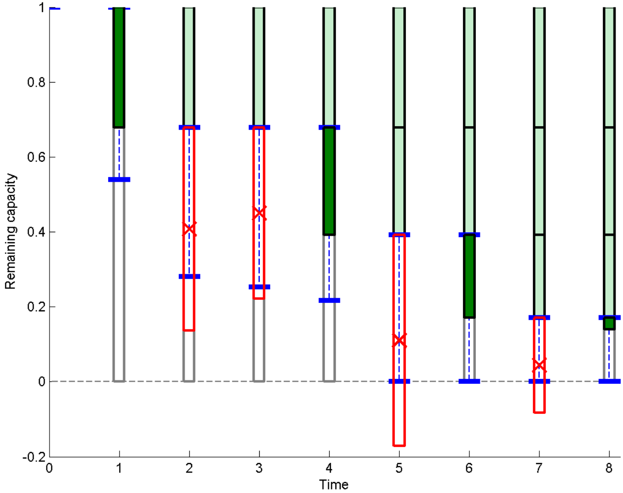
- Knapsack capacity $c = 1$
- Item sizes: $Y_1, Y_2, \ldots, Y_n$ i.i.d. Uniform on $[0, 1]$
- Decision: *include/exclude* Y_t, $1 \leq t \leq n$
- Knapsack policy π_n: the number of items included is

$$N_n(\pi_n) = \max\left\{k : \sum_{i=1}^{k} Y_{\tau_i} \leq 1\right\},$$

  τ_i, index of the ith item included

- Objective: $\sup_{\pi_n} \mathbb{E}\left[N_n(\pi_n)\right]$
- π_n^*: unique optimal Markov deterministic policy based on acceptance intervals

Sequential knapsack problem: dynamics ($n = 8$)



$X_1 = .3213$ $X_2 = 0.5423$ $X_3 = 0.4569$ $X_4 = 0.2865$ $X_5 = 0.5635$ $X_6 = 0.2210$

Sequential knapsack problem: Tight Variance Bounds and a Central Limit Theorem

Theorem (Arlotto, Nguyen, and S. *SPA* 2015)

There is a unique Markov deterministic policy $\pi_n^* \in \Pi(n)$ *such that*

$$\mathbb{E}[N_n(\pi_n^*)] = \sup_{\pi \in \Pi(n)} E[N_n(\pi)]$$

and for such an optimal policy and all $n \geq 1$ *one has*

$$\frac{1}{3}\mathbb{E}[N_n(\pi_n^*)] - 2 \leq \mathrm{Var}[N_n(\pi_n^*)] \leq \frac{1}{3}\mathbb{E}[N_n(\pi_n^*)] + O(\log n)$$

Moreover, one has that

$$\frac{\sqrt{3}\,(N_n(\pi_n^*) - \mathbb{E}[N_n(\pi_n^*)])}{\sqrt{\mathbb{E}[N_n(\pi_n^*)]}} \Longrightarrow N(0,1) \quad \text{as } n \to \infty$$

Observations on the Knapsack CLT

- The central limit theorem holds despite *strong dependence* on the level of remaining capacity and on the time period
- Asymptotic normality tells us almost everything we would like to know!
- Intriguing Open Issue: How is does the central limit theorem for the knapsack depend on the distribution F. This appears to be quite delicate.
- **Major Issue:**

  In what contexts can we get a CLT for the realized reward of an MDP? Is this always an issue of specific problems, or is there some kind of CLT that is relevant to a large class of MDPs. The essence of the matter seems to pivot on the possibility of *useful* CLTs for non-homogenous Markov additive functionals.

1. Some Worrisome Questions with Useful Answers
2. Three motivating examples
3. Rich Class of MDPs where Means Bound Variances
4. Beyond Means-Bound-Variances: Sequential Knapsack Problem
5. **Markov Additive Functionals of Non-Homogenous MCs — with a Twist**
6. Motivating examples: finite-horizon Markov decision problems
7. Dobrushin's CLT and failure of state-space enlargement
8. MDP Focused Central Limit Theorem
9. Conclusions

Markov Additive Functionals — Plus m

Here we are concerned with the possibility of an Asymptotic Gaussian law for partial sums of the form:

$$S_n = \sum_{i=1}^{n} f_{n,i}(X_{n,i}, \ldots, X_{n,i+m}), \qquad \text{for } n \geq 1$$

Framework:

- $\{X_{n,i} : 1 \leq i \leq n+m\}$ are $n+m$ observations from a time *non-homogeneous* Markov chain with general state space $\mathcal{X}$
- $\{f_{n,i} : 1 \leq i \leq n\}$ are real-valued functions on $\mathcal{X}^{1+m}$
- $m \geq 0$: novel twist

Why m?

1. Some Worrisome Questions with Useful Answers
2. Three motivating examples
3. Rich Class of MDPs where Means Bound Variances
4. Beyond Means-Bound-Variances: Sequential Knapsack Problem
5. Markov Additive Functionals of Non-Homogenous MCs — with a Twist
6. Motivating examples: finite-horizon Markov decision problems
7. Dobrushin's CLT and failure of state-space enlargement
8. MDP Focused Central Limit Theorem
9. Conclusions

Dynamic inventory management I

- Model: n periods; i.i.d. demands $D_1, D_2, \ldots, D_n$ uniform on $[0, a]$
- Decision: ordering quantity in each period (instantaneous fulfillment)
- Order up-to functions: if current inventory is x one orders up-to $\gamma_{n,i}(x) \geq x$
- Inventory policy π_n: sequence of order up-to functions
- Time evolution of inventory:

$$X_{n,1} = 0 \qquad \text{and} \qquad X_{n,i+1} = \gamma_{n,i}(X_{n,i}) - D_i \qquad \text{for all } 1 \leq i \leq n$$

- Cost of inventory policy π_n:

$$\mathcal{C}_n(\pi_n) = \sum_{i=1}^{n} \left\{ \underbrace{c(\gamma_{n,i}(X_{n,i}) - X_{n,i})}_{\text{ordering cost}} + \underbrace{c_h \max(0, X_{n,i+1})}_{\text{holding cost}} + \underbrace{c_p \max(0, -X_{n,i+1})}_{\text{penalty cost}} \right\}$$

Special case of the sum S_n with $m = 1$ and one-period cost functions as above

Dynamic inventory management II

- Optimal policy: This is a policy π_n^* such that $\mathbb{E}[\mathcal{C}_n(\pi_n^*)] = \inf_{\pi_n} \mathbb{E}[\mathcal{C}_n(\pi_n)]$ which is based on state and the time dependent order-up-to levels.
- Optimal order-up-to levels (Bulinskaya, 1964):

$$a\left(\frac{c_p - c}{c_h + c_p}\right) = s_1 \le s_2 \le \cdots \le s_n \le a\left(\frac{c_p}{c_h + c_p}\right)$$

such that the optimal order-up-to function

$$\gamma_{n,i}^*(x) = \begin{cases} s_{n-i+1} & \text{if } x \le s_{n-i+1} \\ x & \text{if } x > s_{n-i+1} \end{cases}$$

- What are we after? Asymptotic normality of $\mathcal{C}_n(\pi_n^*)$ as $n \to \infty$ after the usual centering and scaling
  - Uniform demand is not key: unimodal densities with bounded support suffice
  - Instantaneous fulfillment is not key: (some random) lead times can be accommodated (Arlotto and S., 2017)

1. Some Worrisome Questions with Useful Answers
2. Three motivating examples
3. Rich Class of MDPs where Means Bound Variances
4. Beyond Means-Bound-Variances: Sequential Knapsack Problem
5. Markov Additive Functionals of Non-Homogenous MCs — with a Twist
6. Motivating examples: finite-horizon Markov decision problems
7. Dobrushin's CLT and failure of state-space enlargement
8. MDP Focused Central Limit Theorem
9. Conclusions

The minimal ergodic coefficient

- Dobrushin contraction coefficient: Given a Markov transition kernel $K \equiv K(x, dy)$ on $\mathcal{X}$, the Dobrushin contraction coefficient of K is given by

$$\delta(K) = \sup_{x_1, x_2 \in \mathcal{X}} ||K(x_1, \cdot) - K(x_2, \cdot)||_{\mathrm{TV}} = \sup_{\substack{x_1, x_2 \in \mathcal{X} \\ B \in \mathcal{B}(\mathcal{X})}} |K(x_1, B) - K(x_2, B)|$$

- Ergodic coefficient:

$$\alpha(K) = 1 - \delta(K)$$

- Minimal ergodic coefficient: Given a time non-homogeneous Markov chain $\{\widehat{X}_{n,i} : 1 \leq i \leq n\}$ one has the Markov transition kernels

$$K^{(n)}_{i,i+1}(x, B) = P(\widehat{X}_{n,i+1} \in B \,|\, \widehat{X}_{n,i} = x), \qquad 1 \leq i < n,$$

and the minimal ergodic coefficient is given by

$$\alpha_n = \min_{1 \leq i < n} \alpha(K^{(n)}_{i,i+1})$$

A building block: Dobrushin's central limit theorem

Theorem (Dobrushin, 1956; see also Sethuraman and Varadhan, 2005)

For each $n \geq 1$, let $\{\widehat{X}_{n,i} : 1 \leq i \leq n\}$ be n observations of a non-homogeneous Markov chain. If

$$S_n = \sum_{i=1}^{n} f_{n,i}(\widehat{X}_{n,i})$$

and if there are constants $C_1, C_2, \ldots$ such that

$$\max_{1 \leq i \leq n} ||f_{n,i}||_\infty \leq C_n \quad \textit{and} \quad \lim_{n \to \infty} \frac{C_n^2}{\alpha_n^3 \big(\sum_{i=1}^{n} \mathrm{Var}[f_{n,i}(\widehat{X}_{n,i})] \big)} = 0,$$

then one has the convergence in distribution

$$\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}[S_n]}} \Longrightarrow N(0,1), \qquad \textit{as } n \to \infty$$

- Role of the minimal ergodic coefficient α_n

Dobrushin's CLT and Bulinskaya's inventory control problem

- Mean-optimal inventory costs:

$$\mathcal{C}_n(\pi_n^*) = \sum_{i=1}^{n} \Big\{ c(\gamma_{n,i}^*(X_{n,i}) - X_{n,i}) + c_h \max(0, X_{n,i+1}) + c_p \max(0, -X_{n,i+1}) \Big\}$$

- State-space enlargement: define the bivariate chain

$$\{\, \widehat{X}_{n,i} = (X_{n,i}, X_{n,i+1}) : 1 \le i \le n \,\}$$

- Transition kernel of $\widehat{X}_{n,i}$:

$$\begin{aligned} K_{i,i+1}^{(n)}((x,y), B \times B') &= P(X_{n,i+1} \in B, X_{n,i+2} \in B' \,|\, X_{n,i} = x, X_{n,i+1} = y) \\ &= \mathbb{1}(y \in B) P(X_{n,i+2} \in B' \,|\, X_{n,i+1} = y) \end{aligned}$$

- Degeneracy: if $y \in B$ and $y' \in B^c$ one has

$$K_{i,i+1}^{(n)}((x,y), B \times \mathcal{X}) - K_{i,i+1}^{(n)}((x,y'), B \times \mathcal{X}) = 1,$$

so the minimal ergodic coefficient is given by

$$\alpha_n = 1 - \max_{1 \le i < n} \Big\{ \sup_{(x,y),(x',y')} ||K_{i,i+1}^{(n)}((x,y), \cdot\,) - K_{i,i+1}^{(n)}((x',y'), \cdot\,)||_{\mathrm{TV}} \Big\} = 0$$

1. Some Worrisome Questions with Useful Answers
2. Three motivating examples
3. Rich Class of MDPs where Means Bound Variances
4. Beyond Means-Bound-Variances: Sequential Knapsack Problem
5. Markov Additive Functionals of Non-Homogenous MCs — with a Twist
6. Motivating examples: finite-horizon Markov decision problems
7. Dobrushin's CLT and failure of state-space enlargement
8. MDP Focused Central Limit Theorem
9. Conclusions

MDP Focused Markov Additive CLT

Theorem (Arlotto and S., *MOOR*, 2016)

For each $n \geq 1$, let $\{X_{n,i} : 1 \leq i \leq n+m\}$ be $n+m$ observations of a non-homogeneous Markov chain. If

$$S_n = \sum_{i=1}^{n} f_{n,i}(X_{n,i}, \dots, X_{n,i+m})$$

and if there are constants $C_1, C_2, \dots$ such that

$$\max_{1 \leq i \leq n} ||f_{n,i}||_\infty \leq C_n \quad \textit{and} \quad \lim_{n \to \infty} \frac{C_n^2}{\alpha_n^2 \mathrm{Var}[S_n]} = 0,$$

then one has the convergence in distribution

$$\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}[S_n]}} \Longrightarrow N(0,1), \qquad \textit{as } n \to \infty$$

Easy corollary: If the C_n's are uniformly bounded and if the minimal ergodic coefficient α_n is bounded away from zero (i.e. $\alpha_n \geq c > 0$ for all n), then one just needs to show

$$\mathrm{Var}[S_n] \to \infty \qquad \text{as } n \to \infty$$

On $m = 0$ vs $m > 0$

- The asymptotic condition in Dobrushin's CLT imposes a condition on the sum of the individual variances
$$\lim_{n\to\infty} \frac{C_n^2}{\alpha_n^3\big(\sum_{i=1}^n \mathrm{Var}[f_{n,i}(\widehat{X}_{n,i})]\big)} = 0$$
- Our CLT however imposes a condition on the variance of the sum
$$\lim_{n\to\infty} \frac{C_n^2}{\alpha_n^2 \mathrm{Var}[S_n]} = 0$$
- Variance lower bound : when $m = 0$ the two quantities are connected via the inequality
$$\mathrm{Var}[S_n] \geq \frac{\alpha_n}{4} \sum_{i=1}^n \mathrm{Var}[f_{n,i}(X_{n,i})]$$
and our asymptotic condition is weaker
- Counterexample for $m > 0$: when $m > 0$ the variance lower bound fails

Counterexample to variance lower bound when $m = 1$

- Fix $m = 1$ and let $X_{n,1}, X_{n,2}, \ldots, X_{n,n+1}$ be i.i.d. with $0 < \mathrm{Var}[X_{n,1}] < \infty$
- By independence, the minimal ergodic coefficient $\alpha_n = 1$
- For $1 \leq i \leq n$ consider the functions

$$f_{n,i}(x,y) = \begin{cases} x & \text{if } i \text{ is even} \\ -y & \text{if } i \text{ is odd;} \end{cases}$$

and let $S_n = \sum_{i=1}^{n} f_{n,i}(X_{n,i}, X_{n,i+1})$

- For each $n \geq 0$ one has that

$$S_{2n} = 0 \qquad \text{and} \qquad S_{2n+1} = -X_{2n+1,2(n+1)}$$

so the variance

$$\mathrm{Var}[S_n] = O(1) \qquad \text{for all } n \geq 1$$

- On the other hand, the sum of the individual variances

$$\sum_{i=1}^{n} \mathrm{Var}[f_{n,i}(X_{n,i}, X_{n,i+1})] = n\mathrm{Var}[X_{n,1}]$$

MDP Focused CLT Applied to an Inventory Problem

Theorem (Arlotto and S., *MOOR*, 2016)

For each $n \geq 1$, let $\{X_{n,i} : 1 \leq i \leq n+m\}$ be $n+m$ observations of a non-homogeneous Markov chain. If

$$S_n = \sum_{i=1}^{n} f_{n,i}(X_{n,i}, \dots, X_{n,i+m})$$

and if there are constants $C_1, C_2, \dots$ such that

$$\max_{1\leq i\leq n} ||f_{n,i}||_\infty \leq C_n \quad \textit{and} \quad \lim_{n\to\infty} \frac{C_n^2}{\alpha_n^2 \mathrm{Var}[S_n]} = 0,$$

then one has the convergence in distribution

$$\frac{S_n - \mathbb{E}[S_n]}{\sqrt{\mathrm{Var}[S_n]}} \Longrightarrow N(0,1), \qquad \textit{as } n \to \infty$$

Easy corollary: If the C_n's are uniformly bounded and if the minimal ergodic coefficient α_n is bounded away from zero (i.e. $\alpha_n \geq c > 0$ for all n), then one just needs to show

$$\mathrm{Var}[S_n] \to \infty \qquad \text{as } n \to \infty$$

The CLT Applied to the Inventory Management Problem

Dynamic Inventory Management: It is always optimal to order a positive quantity if the inventory on-hand drops below a level $s_1 > 0$

- Minimal ergodic coefficient lower bound: $\alpha_n \geq \dfrac{s_1}{a}$ for all $n \geq 1$
- Variance lower bound: $\mathrm{Var}[\mathcal{C}_n(\pi_n^*)] \geq K(s_1)n$
- Asymptotic Gaussian law: as $n \to \infty$

$$\frac{\mathcal{C}_n(\pi_n^*) - \mathbb{E}[\mathcal{C}_n(\pi_n^*)]}{\sqrt{\mathrm{Var}[\mathcal{C}_n(\pi_n^*)]}} \Longrightarrow N(0,1)$$

1. Some Worrisome Questions with Useful Answers
2. Three motivating examples
3. Rich Class of MDPs where Means Bound Variances
4. Beyond Means-Bound-Variances: Sequential Knapsack Problem
5. Markov Additive Functionals of Non-Homogenous MCs — with a Twist
6. Motivating examples: finite-horizon Markov decision problems
7. Dobrushin's CLT and failure of state-space enlargement
8. MDP Focused Central Limit Theorem
9. Conclusions

Summary

- **Main Message:** One need not feel *too* guilty applying mean-focused MDP tools. We know they seem to work in practice, and there is a growing body of knowledge that helps to explain why they work.

- **First:** In a rich class of problems, there are *a priori* bounds on the variance that are given in terms of the mean reward and the bound on the individual rewards.

  Three simple properties characterize this class.

- **Second:** When more information is available on the Markov chain of decision states and post-decision reward functions, one has good prospects for a Central Limit Theorem.

  Application of the MDP focused CLT is not work-free. Nevertheless, because of the MDP focus, one has a much shorter path to a CLT than one could realistically expect otherwise. At a minimum, we know reasonably succinct sufficient conditions for a CLT.

Thank you!

Bounding the variance by the mean: proof details

- The martingale difference

$$d_t = M_t - M_{t-1} = r(t, X_t, Y_t, A_t^*) + v_{t+1}(X_{t+1}) - v_t(X_t)$$

- Add and subtract $v_{t+1}(X_t)$ to obtain

$$\begin{aligned} d_t =& \ v_{t+1}(X_t) - v_t(X_t) \\ &+ \ r(t, X_t, Y_t, A_t^*) + v_{t+1}(X_{t+1}^*) - v_{t+1}(X_t) \end{aligned}$$

- Recall: X_t is $\mathcal{F}_{t-1}$-measurable
- Hence, α_t is $\mathcal{F}_{t-1}$-measurable and $\alpha_t = -\mathbb{E}[\beta_t \,|\, \mathcal{F}_{t-1}]$, so

$$\mathbb{E}[d_t^2 \,|\, \mathcal{F}_{t-1}] = \mathbb{E}[\beta_t^2 \,|\, \mathcal{F}_{t-1}] + 2\alpha_t \mathbb{E}[\beta_t \,|\, \mathcal{F}_{t-1}] + \alpha_t^2 = \mathbb{E}[\beta_t^2 \,|\, \mathcal{F}_{t-1}] - \alpha_t^2$$

- Since $\alpha_t^2 \geq 0$ and $\beta_t \leq r(t, X_t, Y_t, A_t^*)$

$$\mathbb{E}[d_t^2 \,|\, \mathcal{F}_{t-1}] \leq \mathbb{E}[\beta_t^2 \,|\, \mathcal{F}_{t-1}] \leq K \, \mathbb{E}[r(t, X_t, Y_t, A_t^*) \,|\, \mathcal{F}_{t-1}]$$

◂ Back to Proof Sketch

Bounding the variance by the mean: proof sketch

- For $0 \leq t \leq n$, the process
$$M_t = R_t(\pi_n^*) + v_{t+1}(X_{t+1})$$
is a martingale with respect to the natural filtration $\mathcal{F}_t = \sigma\{Y_1, \ldots, Y_t\}$
- $M_0 = \mathbb{E}[R_n(\pi_n^*)]$ and $M_n = R_n(\pi_n^*)$
- For $d_t = M_t - M_{t-1}$,
$$\mathrm{Var}[M_n] = \mathrm{Var}\left[R_n(\pi_n^*)\right] = \mathbb{E}\left[\sum_{t=1}^{n} d_t^2\right]$$
- "Some rearranging" and an application of reward non-negativity and boundedness, existence of a do-nothing action, and optimal action monotonicity gives
$$\mathbb{E}[d_t^2 \mid \mathcal{F}_{t-1}] \leq K\,\mathbb{E}[r(t, X_t, Y_t, A_t^*) \mid \mathcal{F}_{t-1}]$$
- Taking total expectations and summing gives
$$\mathrm{Var}\left[R_n(\pi_n^*)\right] \leq K\,\mathbb{E}\left[R_n(\pi_n^*)\right]$$
- Crucial here: $X_{t+1} = f(t, X_t, Y_t, A_t)$ is $\mathcal{F}_t$-measurable! □

Bounding the variance by the mean: proof details

- The martingale difference

$$d_t = M_t - M_{t-1} = r(t, X_t, Y_t, A_t^*) + v_{t+1}(X_{t+1}) - v_t(X_t)$$

- Add and subtract $v_{t+1}(X_t)$ to obtain

$$\begin{aligned} d_t = & \overbrace{v_{t+1}(X_t) - v_t(X_t)}^{\alpha_t} \\ & + \underbrace{r(t, X_t, Y_t, A_t^*) + v_{t+1}(X_{t+1}) - v_{t+1}(X_t)}_{\beta_t} \end{aligned}$$

- Recall: X_t is $\mathcal{F}_{t-1}$-measurable

- Hence, α_t is $\mathcal{F}_{t-1}$-measurable and $\alpha_t = -\mathbb{E}[\beta_t \mid \mathcal{F}_{t-1}]$, so

$$\mathbb{E}[d_t^2 \mid \mathcal{F}_{t-1}] = \mathbb{E}[\beta_t^2 \mid \mathcal{F}_{t-1}] + 2\alpha_t \mathbb{E}[\beta_t \mid \mathcal{F}_{t-1}] + \alpha_t^2 = \mathbb{E}[\beta_t^2 \mid \mathcal{F}_{t-1}] - \alpha_t^2$$

- Since $\alpha_t^2 \geq 0$ and $\beta_t \leq r(t, X_t, Y_t, A_t^*)$

$$\mathbb{E}[d_t^2 \mid \mathcal{F}_{t-1}] \leq \mathbb{E}[\beta_t^2 \mid \mathcal{F}_{t-1}] \leq K\, \mathbb{E}[r(t, X_t, Y_t, A_t^*) \mid \mathcal{F}_{t-1}]$$

Remark (Uniform Boundedness)

The Variance Bound still holds if there is a constant $K < \infty$ such that

$$\mathbb{E}[r^2(t, X_t, Y_t, A_t^*)] \leq K\mathbb{E}[r(t, X_t, Y_t, A_t^*)]$$

uniformly in t.

Remark (Bounded Differences Martingale)

The Bellman martingale has bounded differences. In fact we also have that

$$|d_t| = |M_t - M_{t-1}| = |\alpha_t + \beta_t| \leq K$$

for all $1 \leq t \leq n$*.*

References I

Alessandro Arlotto and J. Michael Steele. Optimal sequential selection of a unimodal subsequence of a random sequence. *Combinatorics, Probability and Computing*, 20 (06):799–814, 2011. doi: 10.1017/S0963548311000411.

Carri W. Chan and Vivek F. Farias. Stochastic depletion problems: effective myopic policies for a class of dynamic optimization problems. *Math. Oper. Res.*, 34(2): 333–350, 2009. ISSN 0364-765X. doi: 10.1287/moor.1080.0364. URL `http://dx.doi.org/10.1287/moor.1080.0364`.

E. G. Coffman, Jr., L. Flatto, and R. R. Weber. Optimal selection of stochastic intervals under a sum constraint. *Adv. in Appl. Probab.*, 19(2):454–473, 1987. ISSN 0001-8678. doi: 10.2307/1427427.

C. Derman, G. J. Lieberman, and S. M. Ross. A stochastic sequential allocation model. *Operations Res.*, 23(6):1120–1130, 1975. ISSN 0030-364X.

Jason D. Papastavrou, Srikanth Rajagopalan, and Anton J. Kleywegt. The dynamic and stochastic knapsack problem with deadlines. *Management Science*, 42(12):1706–1718, 1996. ISSN 00251909.

Gregory P. Prastacos. Optimal sequential investment decisions under conditions of uncertainty. *Management Science*, 29(1):118–134, 1983. ISSN 00251909.

References II

Paul A. Samuelson. Lifetime portfolio selection by dynamic stochastic programming. *The Review of Economics and Statistics*, 51(3):239–246, 1969. ISSN 00346535.

Kalyan T. Talluri and Garrett J. van Ryzin. *The theory and practice of revenue management*. International Series in Operations Research & Management Science, 68. Kluwer Academic Publishers, Boston, MA, 2004. ISBN 1-4020-7701-7.